

Data Enhancement — Enrichment Maturity Assessment

Please select exactly one answer per question. Select the option that best describes your current situation today, not the situation you are working toward.

Dimension 8 — Data Enhancement — Enrichment

Definition*: Data sources enriched, for example, with annotations, image labels, phenotype definitions, derived variables, and labels produced by natural language processing. A derived field is a value your organization calculates or assigns; it is not recorded directly at the source.

1. Does your data contain derived fields or added annotations?*

- | | |
|---|--------|
| ☐ No. The data contains only what was recorded at source. | 1 pt. |
| ☐ Yes, some derived fields or annotations that we created ourselves. | 2 pts. |
| ☐ Yes, and they are also used by other data sources or organizations. | 3 pts. |
| ☐ Yes, and they are produced by, or used in, a peer-reviewed algorithm, one published and assessed by independent experts. | 4 pts. |
| ☐ Yes, and they come from a national or international report. | 5 pts. |

You can support this answer with: your list of derived fields with their definitions.

2. How widely are derived fields used across your data?*

- | | |
|--|--------|
| ☐ Nowhere. | 1 pt. |
| ☐ A few fields in one dataset. | 2 pts. |
| ☐ Across our main datasets. | 3 pts. |
| ☐ Across all datasets offered for secondary use. | 4 pts. |
| ☐ Across all datasets, generated and refreshed automatically. | 5 pts. |

You can support this answer with: your data dictionary, marking which fields are derived.

3. Who uses the enriched data?*

- | | |
|--|--------|
| ☐ Nobody uses it. | 1 pt. |
| ☐ Our own teams only. | 2 pts. |
| ☐ Our own teams and named partners. | 3 pts. |
| ☐ External researchers, routinely. | 4 pts. |
| ☐ External researchers, routinely, and it is cited in publications or official reports. | 5 pts. |

You can support this answer with: project list, publication list referencing your derived variables.

4. What are the derived fields based on?*

- | | |
|--|--------|
| ☐ No defined basis, so they were created ad hoc. | 1 pt. |
| ☐ An internal rule that is written down. | 2 pts. |
| ☐ A national guideline, or an algorithm validated locally. | 3 pts. |
| ☐ A peer-reviewed published algorithm. | 4 pts. |
| ☐ A national or international standard, or a definition taken from an international report. | 5 pts. |

You can support this answer with: the reference or citation for each derivation rule.

5. Can users tell derived fields apart from source fields?*

- | | |
|---|--------|
| ☐ No. Derived and source fields are indistinguishable. | 1 pt. |
| ☐ It is mentioned informally when we hand over data. | 2 pts. |
| ☐ Derived fields are marked and described in the dataset documentation. | 3 pts. |
| ☐ The derivation rule and its source are documented for every derived field. | 4 pts. |
| ☐ As above, machine-readable and published together with the dataset. | 5 pts. |

You can support this answer with: your data dictionary entry for a derived field.

YOUR NOTES FOR THIS DIMENSION

What sits behind these answers: the practices and tools you rely on, recent improvements, and the gaps that explain the lower scores. This is the material for the supporting narrative on the Overall Score sheet.

DIMENSION SCORE

Add up the points of all 5 answers and divide by 5. Keep one decimal.

YOUR MATURITY LEVEL

Round the dimension score down to the nearest whole number.

Level 1 Initial Starting out · 1 Point	Level 2 Emerging Written down · 2 Points	Level 3 Defined In practice · 3 Points	Level 4 Managed Measured · 4 Points	Level 5 Optimized Improving · 5 Points
--	--	--	---	--

The five maturity levels in this dimension

Your five answers average to one level. Each row describes what that level typically looks like across all five questions, so you may be ahead of it on some points and behind on others.

Level 1 Initial	No derived fields or annotations, so nothing to use, document, or distinguish.
Level 2 Emerging	A few self-made derived fields in one dataset, used internally, mentioned informally.
Level 3 Defined	Derived fields across the main datasets, shared with partners, described in the documentation.
Level 4 Managed	Peer-reviewed derivations across all shared datasets, used by external researchers, documented.
Level 5 Optimized	Standard or report-based definitions, generated automatically, cited, published machine-readable.

Your next step: find the row for the level you reached, then read the row directly below it. That row is what you need to achieve to move up one level in this dimension.

Note: Every question in this dimension uses the same 1–5 scale, and the same real-world situation is worth the same number of points in every question, so both the lowest and the highest level are reachable. Transfer your maturity level for this dimension to the *Maturity Overall Score Calculation* sheet, which brings all ten dimensions together.

References*

Question 1 follows the maturity levels of this dimension in the QUANTUM data holders' maturity model (Deliverable 1.2 v3.5); the wording is EHDS4ALL's own. Underlying source given there for this dimension: Adapted HDRUK. Questions 2 to 5 are EHDS4ALL, developed from the WP3 workshop of August 2024 and the expert interviews of October 2024. The five level statements below combine all five questions, because the dimension score is the average of all five answers.

Jones, M., & Bernal-Delgado, E. (2025). **Deliverable 1.2 — Specification for the assessment of data holders' maturity** (Version 3.5.0). QUANTUM project. Zenodo. doi: 10.5281/zenodo.14944767

Gordon, B., Barrett, J., Fennessy, C., et al. (2021). Development of a data utility framework to support effective health data curation. *BMJ Health & Care Informatics*, 28(1), e100303.

EHDS4ALL (2024). Protocol, Workshop Kick-off WP3, Cologne, 19 August 2024. Work Package 3, dimension validation and maturity levels.

EHDS4ALL (2024). Expert interviews with health data holders and maturity model developers, October 2024. Conducted by inav and the University of Cologne; participants anonymized.